

1 1. A computer-implemented method for searching a large number of
2 hypertext documents in accordance with a current search query, comprising:
3 determining which of the hypertext documents are expert documents;
4 ranking the expert documents in accordance with the current search
5 query by:
6 determining a level score for each of the expert documents;
7 determining a fullness factor for each key phrase on each of the
8 expert documents; and
9 determining an expert score for each expert document in
10 accordance with the level score of the expert document and
11 the fullness factors for the key phrases of the expert
12 document;
13 ranking target documents pointed to by the ranked expert documents;
14 and
15 returning a results list based on the ranked target documents.

1 2. The computer-implemented method of claim 1, wherein the hypertext
2 documents are pages in the world wide web.

1 3. The computer-implemented method of claim 1, wherein the hypertext
2 documents are sites in the world wide web.

1 4. The computer-implemented method of claim 1, wherein the hypertext
2 documents are documents in a hypertext database.

1 5. The computer-implemented method of claim 1, wherein an expert
2 reverse index is constructed in memory for keywords appearing in the expert
3 documents, the expert reverse index identifying the location of the keywords in
4 the expert documents.

1 6. The computer-implemented method of claim 5, wherein a keyword of
2 an expert document is included in the expert reverse index if the keyword is
3 part of a key phrase that qualifies at least one URL in the expert document.

1 7. The computer-implemented method of claim 6, wherein a key phrase
2 qualifies a URL if the URL is within the scope of the key phrase in the expert
3 document.

1 8. The computer-implemented method of claim 6, wherein a key phrase
2 in an HTML title qualifies all URLs in the entire document.

1 9. The computer-implemented method of claim 6, wherein a key phrase
2 in an HTML heading qualifies all URLs in that portion of the document before a
3 next HTML heading in the document of greater or equal importance.

1 10. The computer-implemented method of claim 6, wherein a key phrase
2 in an HTML anchor qualifies the URLs in the anchor.

(b) (1)
11. The computer-implemented method of claim 1, wherein determining
2 which of the hypertext documents are expert documents includes:

3 determining a document having at least a predetermined number of
4 outlinks to be an expert document if the document also points to at
5 least the predetermined number of targets on distinct non-affiliated
6 hosts.

12. The computer-implemented method of claim 11, wherein expert
2 documents additionally must point to documents that share the same broad
3 classification.

14. The computer-implemented method of claim 1, wherein ranking
2 target documents pointed to by the expert documents includes:

3 determining a plurality of edge scores for each target document, where
4 an edge score is determined for edges between the expert documents
5 and the target document;

6 determining a target score in accordance with the edge scores of the
7 target document;

8 ranking the target documents in accordance with the target scores.

15. The computer-implemented method of claim 14, further including:
2 determining an edge score only for those links to the target document
3 from a predetermined number of top-ranked expert documents.

16. The computer-implemented method of claim 14, further including
1 selecting target documents to be ranked that are linked to by at least two
2 mutually non-affiliated selected expert documents, where the selected target
3 also is not affiliated with the expert documents.

17. The computer-implemented method of claim 14, where an edge score
2 between an expert document and a target document $ES(E,T)$ is determined as
3 follows, where ExpertScore reflects the rankings of the expert documents:
4 a) find #occurrences of each keyword in all key phrases of expert
5 document E
6 b) if the #occurrences for any keyword in E is 0: $ES(E,T)=0$
7 else $ES(E,T)=\text{ExpertScore}(E) * \text{sum of #occurrences for all keywords.}$

18. The computer-implemented method of claim 14, wherein, if two
2 affiliated experts have edges to the same target, the edge having a lower edge
3 score is discarded and is not used to determine the target score.

19. The computer-implemented method of claim 1, wherein two
2 hypertext documents are affiliated if at least one of the following is true: 1) they
3 share the same rightmost non-generic suffix and 2) they have an IP address in
4 common.

20. An apparatus that searches a large number of hypertext documents in
2 accordance with a current search query, comprising:

3 a software portion configured to determine which of the hypertext
4 documents are expert documents;
5 a software portion configured to rank the expert documents in accordance
6 with the current search query by:
7 determining a level score for each of the expert documents;
8 determining a fullness factor for each key phrase on each of the
9 expert documents; and
10 determining an expert score for each expert document in
11 accordance with the level score of the expert document and
12 the fullness factors for the key phrases of the expert
13 document;
14 a software portion configured to rank target documents pointed to by the
15 ranked expert documents; and
16 a software portion configured to return a results list based on the ranked
17 target documents.

1 21. A computer program product, comprising:
2 a computer usable medium having computer readable instructions stored
3 therein to search a large number of hypertext documents in accordance
4 with a current search query, including:
5 computer readable program code devices for causing a computer to
6 determine which of the hypertext documents are expert documents;
7 computer readable program code devices for causing a computer to rank
8 the expert documents in accordance with the current search query by:
9 determining a level score for each of the expert documents;
10 determining a fullness factor for each key phrase on each of the
11 expert documents; and

12 determining an expert score for each expert document in
13 accordance with the level score of the expert document and
14 the fullness factors for the key phrases of the expert
15 document;
16 computer readable program code devices for causing a computer to rank
17 target documents pointed to by the ranked expert documents; and
18 computer readable program code devices for causing a computer to return
19 a results list based on the ranked target documents.
